1994 SIGAPL Survey Results

Dick Holt
SIGAPL Member-at-Large
3802 N. Richmond Street
Arlington, VA 22207 USA
Tel: 703-528-7624 home, 202-586-4449 office
Fax and BBS\APL: 703-528-7617, 1200-14440b, 24-hr
E-mail: dick.holt@acm.org

I. Summary

The purpose of this Survey is to provide current world-wide information about APL users, and to identify how SIGAPL can better serve its customers. Three data sets describe the main results: (1) the full survey sample, (2) the SIG subset ((belong to SIGAPL) *or* (read *APL Quote Quad* regularly)), and (3) the Non-SIG subset (*not* SIG). The split sample improves survey representativeness and sharpens comparisons of interest to SIGAPL.

Is the survey a representative cross-section of SIGAPL? There's multiple evidence that it is. You're now invited to explore what 13% of all SIGAPL members and 84 others had to say. Here are highlights of the full Survey:

What features of *APL Quote Quad* (QQ) are most and least useful?

Most Useful	Least Useful
Conference Proceedings	Windows
Algorithms	Business
New product reviews	ISO Standards

What APL information sources are most and least useful?

Most Useful	Least Useful
Vector	Big APL
Quote Quad	BBS\APL
comp.lang.apl	Education Vector

What SIGAPL Conference features are most and least satisfactory?

Most Satisfactory	Least Satisfactory
Proceedings	Software Exchange
Location	Cost
Vendor Information	Business/Job opportunities

44% of survey respondents don't read QQ regularly. Why?

17% Don't know what QQ is, or how to get it.
11% Dissatisfied with QQ content or ACM subscription fulfillment.
9% QQ is hard to get (libraries, some countries).
9% Newcomer to APL or J.

49% of respondents haven't attended a SIGAPL conference recently. Why?

18% High costs, or costs not paid by employer.
9% Conferences aren't sufficiently work or business related.
9% Don't have time.
8% Newcomer to APL or J, or didn't know about conferences.

44% responded to the question *"How can SIGAPL better meet your needs?"*

10% Provide more job/business related information.
9% Facilitate wider connection of APL to other languages.
8% Improve content of conferences and publications.

II. Is the Survey a Representative Cross-Section of SIGAPL?

To answer this question, we compare survey responses with ACM demographic data for SIGAPL (Ref 1). If you aren't interested in statistical details, skip ahead to Section III.

25 US states accounted for 64% of all responses, and 10% came from at least 6 Canadian Provinces. 5% were from the UK, and 12% from 9 other European nations. Others were from Australia, Brazil, South Africa, Turkey, Japan, New Zealand, and Singapore. We asked for, but didn't get, responses from Israel, Russia, Portugal, Hong Kong, and West Virginia.

Comparing survey responses at a highly-aggregated level:

Continent	Survey %	SIG %	ACM %	Sigma	Non-Sig %
North America	74	72	69	4.6	78
Europe	17	20	23	4.4	11
Other	9	8	8	2.8	11
N	186	121	65	783	65

Column 1 is data from all 186 respondents. Column 2 is the SIG subset. Column 3 is ACM data for SIGAPL. Column 5 is the Non-SIG subset. Column 4 needs some technical explanation. It's a bootstrap estimate of the standard deviations for ACM geographic areas. *"Bootstrap"* is a statistician's term of art for repeated random sampling, with replacement, from the same data set. Bootstrap estimates above are based on 5000 random samples of size 121 drawn from a set of 783 letters (121?783), of which 69% are *"N"*, 23% *"E"*, and 8% *"O"* — i.e. the geographically partitioned ACM data (for North America, Europe, and Other). In each sample, the number of Ns, Es, and Os vary randomly about their means, thus enabling estimates of their standard deviations (Ref 2). If we (mentally) conducted the survey 5000 times, we would expect variations due to chance alone to be within ±2 sigma in 95% of the cases. If we had used 186?783 for the full sample, sigma would be about 25% larger ((186 *divided_by* 121)*0.5).

Disaggregating to areas having at least 12 SIGAPL members (arbitrary):

Nations	Survey %	SIG %	ACM %	Sigma
US	64.5	62.0	59.1	5.0
Canada	9.7	9.9	9.1	2.9
Germany	3.8	5.0	5.6	2.3
UK	5.4	5.8	2.7	1.7
Japan	0.5	1.0	2.7	1.7
Italy	0.5	0.8	2.4	1.6
Sweden	1.6	2.5	2.2	1.5
France	1.6	1.7	2.2	1.5
Australia	1.6	2.5	2.0	1.3
Netherlands	1.1	1.7	1.8	1.3
Belgium	1.1	0.8	1.8	1.3
Denmark	0.5	0.8	1.8	1.3
Switzerland	0.5	0.8	1.8	1.3
Finland	0.5	0.0	1.7	1.3
Regions: US				
New York	11.3	15.7	11.7	3.3
California	9.7	10.7	9.6	3.0
New Jersey	1.6	1.6	4.5	2.1
Massachusetts	2.1	3.3	4.3	2.1
Texas	6.5	4.1	4.1	2.0
Maryland	10.7	8.3	3.3	1.8
Pennsylvania	5.4	4.1	2.7	1.6
Illinois	0.5	0.8	2.4	1.6
Ohio	1.1	1.7	1.5	1.2
Connecticut	1.1	1.6	1.5	1.2
Regions: Canada				
Ontario	4.3	5.7	4.2	2.0
Quebec	0.5	0.0	1.7	1.3

▸ ***Is There Sampling Bias?***

With one exception, the full survey and the SIG subset are within ±2 sigma confidence limits of ACM data at continental, national, and regional levels. The exception? Maryland is over-represented. Sampling bias exists if Marylanders are significantly different than non-Marylanders. We investigate a split sample:

	Non-Marylanders	Marylanders
Member of SIGAPL	56%	45%
Read QQ regularly	58%	45%
QQ features ranked	37%	23%
Attend Conference recently	54%	25%
Conference features ranked	30%	11%
Information sources ranked	25%	23%

A chi-square test (Ref. 3) shows that none of these differences are significant (95%). A factor that probably is significant: it was easy for me to distribute the survey via the Potomac Chapter of SIGAPL. Data from six fewer Marylanders would bring the full sample within two sigma. Four fewer would suffice for the SIG subset. A different example of over-response occurred for the New York Tool of Thought Conference (see below).

Beyond Maryland, only those who got the survey could respond to it, and not everybody did. It was distributed via:

- `comp.lang.apl` and two other Internet news groups
- The *Swansea*, *APL94*, *Tools of Thought*, and *APLTex* Conferences
- *Vector*, *Gimme Arrays!*, *the APL perspective*, *Les Nouvelles d'APL* (translated into French), and the *APL BUG*
- Many APL user groups world-wide
- Renaissance Data Systems, a source of APL books
- On disk at the APL94 Software Exchange
- The APL *"White Pages"* on the Internet
- Two BBSs specializing in APL and one BBS for actuaries
- Fax, e-mail, and diskette to my own contact lists
- Manugistics posted it on their BBS
- IBM posted it on an internal forum.

SIGAPL thanks all who helped to distribute this survey, et à Les Nouvelles d'APL, *notre remerciements*.

▸ ***Is There Non-Response Bias From Conferences?***

The New York *"Tool of Thought"* (ToT) Conference posed a novel question of response bias. A copy of APL2 for OS/2 was given as a prize in a drawing to those who completed the survey, resulting in near 100% response. Did this bias the sample compared to sources offering no incentive? It didn't. Analyses similar to those for Maryland showed that ToT data weren't detectably biased. ToT data may have actually improved the demographic spread of the survey; about 60% of ToT respondents weren't from New York.

In contrast, there was widespread non-response from other conferences in 1994. Either survey forms weren't well distributed, or 98% of attenders didn't respond, or both.

Conference	Attenders (approximate)	Responses
Swansea - July	40	1
APL94 - September	175	3
APLTex - October	30	0

Many of these conference attenders may have responded by other means. 13% of all SIGAPL members responded to the survey in one way or another. What probably saved the bacon was that 66% of all respondents use an average of 3.3 sources of APL information, and 80% of those attending a SIGAPL conference in the past 5 years use an average of 4.2. As a result, parallelism in Survey distribution enabled conference attenders to respond via multiple channels — which apparently they did.

For Swansea and APLTex: the UK and Texas are somewhat over-represented in the Survey compared to ACM data, but not significantly so. For APL94: Europe (except the UK and Sweden) is somewhat under-represented compared to ACM data, but not significantly so. My best guess? Low returns from these conferences didn't much affect the question of representativeness.

In summary, full survey data and the SIG subset portray SIGAPL with good statistical accuracy. *"Good"* means: (1) survey data are statistically comparable to ACM demographic data for SIGAPL with the exception of Maryland, and (2) overall SIGAPL response rate was high (13% of all SIGAPL members), despite low returns from 3 conferences.

III. Respondent Characteristics

- ***Years using APL?***

The distribution of respondent's *"years using APL"* shows a peak at 16-20 years, and is skewed toward *"years"* greater than the mean.

Years using (gt-le)	Percent of Respondents
0-5	12
6-10	11
11-15	22
16-20	31
21-25	15
26-30	6
31-35	1

Experience Parameters	
Mean	15.6 years
Median	17.5
Mode	20
Standard dev.	7.6
N (6 blanks)	180

12% of the survey respondents are relative newcomers. This is roughly consistent with a steady-state replenishment rate in a population of people with 40 year careers.

- ***Hardware/Software most used in your work?***

Survey respondents use a wide range of hardware and software.

Hardware Used	Software: Percentage Used								
	APL* Plus	IBM APL2	Dya-log	J	Iver-son	APL 68000	Sharp	Other	Total
PC	24.7	10.8	6.5	5.4	1.6	1.1	0.5	5.4	55.9
Mac	0	0	0	0	0	4.3	0	1.6	5.9
Mini/WS	0	2.2	4.8	1.6	0	0	0	4.3	12.9
Mainframe	2.7	11.8	1.1	0	0	0	3.2	2.2	21.0
Blank	1.1	1.6	0	1.1	0	0	0.5	0	4.3
Total	28.5	26.3	12.4	8.1	1.6	5.4	4.3	13.4	100.0

"Other" software includes some non-APL (e.g. MatLab, WP), a few other APLs (e.g. APL90), and blanks. For PC users, DOS, Windows, and OS/2 are used by 52%, 37%, and 15% of respondents respectively. Lacking vendor, piracy, and public domain data for comparison, this table is useless (but interesting).

- ***J***

Again, a split sample:

	J	not J
SIGAPL Member	85%	51%
Read QQ regularly	71%	51%
Attended APL Conference recently	54%	47%
Ranked c.l.a.	77%	40%

Respondents answering *"J"* to the question *"what software do you use most in your work?"* aren't significantly different from those not answering *"J"* (chi-square, 95%). The 13 J respondents are numerically comparable to the 10 Marylanders in the SIG subset.

IV. Usefulness of Quote Quad Features

Question C asked: *"How useful to you are the QQ features below? (1=most; 5=least)."* In the `16 x 5` matrix `M` below, `M[2;3]=22` is the number of respondents who rated *"Algorithms"* as *"3"* on the 1-5 scale.

Table 1: Usefulness of QQ Features

Full Sample N	%	QQ Feature	most - useful - least 1	2	3	4	5	ARS
80	43	Conference Proceedings	43	21	11	1	4	1.78
88	47	Algorithms	40	20	22	5	1	1.94
79	42	Product reviews	36	21	14	5	3	1.96
80	43	Scientific	24	24	22	7	3	2.26
70	38	Education	21	19	19	7	4	2.34
60	32	Bibliographies	18	16	15	7	4	2.38
64	34	Letters to the Editor	15	20	21	5	3	2.39
71	38	Interviews	10	28	25	4	4	2.49
69	37	Other languages (e.g. J)	20	19	9	7	14	2.65
50	27	Telecommunications	10	10	19	9	2	2.66
68	37	Windows	10	20	20	8	10	2.82
55	30	Frequency of publication	9	12	18	11	5	2.84*
64	34	Business	13	14	16	11	10	2.86
56	30	ISO Standards	5	16	23	3	9	2.91
54	29	Timeliness of articles	7	8	25	7	7	2.98*
44	24	Bilingual articles	3	4	8	3	26	4.02*

* Data problems: a reviewer noted that questions on timeliness of QQ articles and frequency of publication are ambiguous. Respondents may have referred to current issues, or they may reflect earlier times when QQ wasn't regularly published — a situation that was fixed by the 1993 SIGAPL election. Ranking of Bilingual articles is inexplicable. QQ has never published a bilingual article. Despite extensive review by the SIGAPL Executive Committee, these three questions were imprecise, and we ignore their data.

Tables are sorted by the *Average Rank Score* (ARS): `(M x ((1 rho rho M),5) rho iota 5) divided_by +/M`, where `M` is the `((1 rho rho M),5)` matrix of rank scores. This is similar to the approach of Ref. 4, adjusted for the different numbers of people who responded to different items.

A guideline for statistical significance: large differences in ARS imply that rank preferences don't arise from chance alone. Large differences are significant, small differences aren't. If a particular comparison is interesting to you, just do a quick test for significance.

Measures other than ARS are possible. A *"net utility measure"*: `(+/M[;1 2]-M[;4 5]) divided_by +/M[;1 2 4 5]` sharpens the results by omitting the indifferent ratings of *"3"* (Ref 5). This measure gives slightly different results, at the expense of not using all the data.

A split sample table for QQ features is unnecessary. It's identical to Table 1 for the SIG subset, and is all zeros for the Non-SIG subset.

V. Usefulness of APL Information Sources

Editors, conference organizers, vendors, educators, actuaries, scientists, financial analysts, all have different interests. Hundreds of interesting comparisons are possible. There's no space for them here. You can do it yourself for areas that interest you most. Data problems are marked "*". Further, you may compare ARS data between the full sample and the split sample tables, and between the SIG and non-SIG subsets. All tables (except Table 4) are constructed exactly the same way.

Question D asked: ***"How useful to you are the sources of APL information below? (1=most; 5=least)."***

Table 2: Usefulness of APL Information Sources

Full Sample N	%	Information Source	most 1	2	3	4	least 5	ARS
97	52	Vector	63	18	11	5	0	1.57
104	56	Quote Quad	42	29	24	5	4	2.04
81	44	comp.lang.apl	35	23	10	8	5	2.07
37	20	the APL perspective	17	9	5	0	6	2.16
39	21	Gimme Arrays!	12	16	5	2	4	2.23
51	27	Vendors	15	11	15	4	6	2.51
42	23	APL News	11	9	10	6	6	2.69*
22	12	APL BUG	5	7	2	4	4	2.77
25	13	Big APL	6	8	1	4	6	2.84
26	14	BBS\APL	8	4	3	5	6	2.88
36	19	Education Vector	6	8	11	5	6	2.92
19	10	APL CAM	5	2	3	3	6	3.16*
14	8	Les Nouvelles d'APL	4	1	1	1	7	3.43*

*** Data problems: a reviewer noted** ***"the survey catered primarily to English-speaking readers, so the results may not do justice to non-English publications."*** **He's right. Also, APL News has folded, so time-lag may be a problem. We ignore these data.**

A different picture emerges from the split sample.

Table 3: Split Sample: APL Information Sources

SIG N	%		ARS	Non-SIG N	%		ARS
83	69	VECTOR	1.59	14	22	VECTOR	1.43
56	46	comp.lang.apl	1.84	7	11	APL News	2.43*
96	79	Quote Quad	1.97	25	38	comp.lang.apl	2.60
31	26	the APL perspective	2.00	3	5	APL BUG	2.67
34	28	Gimme Arrays!	2.15	5	8	Gimme Arrays!	2.80
42	35	Vendors	2.38	8	12	Quote Quad	2.88
18	15	BBS\APL	2.72	2	3	APL CAM	3.00
35	29	APL News	2.74*	2	3	Les Nouvelles d'APL	3.00*
22	18	Big APL	2.77	6	9	the APL perspective	3.00*
19	16	APL BUG	2.79	9	14	Vendors	3.11
33	27	Education Vector	2.85	8	12	BBS\APL	3.25
17	14	APL CAM	3.18*	3	5	Big APL	3.33
12	10	Les Nouvelles d'APL	3.50*	3	5	Education Vector	3.67

An interesting picture emerges from a pairwise comparison of information sources. In Table 4, `M[j;k]` is the rank of the *j*th source, as rated by those in column k who ranked *both* source j and source k. Data in Table 4 are the 145 respondents who ranked at least one source of information, scrambled into 156 very different subsets compared to Table 2.

Table 4: A Pairwise Rank Kaleidoscope

	VEC	QQ	cla	Per	Gim	Ven	* New	BUG	Big	BBS	EdV	* CAM	* Les	row sums
Vector	1	1	1	1	1	1	1	1	1	1	1	1	1	13
QQ	3	3	3	4	2	2	2	3	2	2	2	2	2	32
c.l.a	2	4	2	3	3	4	3	2	4	3	3	3	4	40
APL persp	4	2	4	2	4	3	4	4	3	4	5	4	3	46
Gimme Arr	5	5	5	6	5	5	5	5	6	11	4	5	5	72
Vendors	6	6	6	5	6	6	7	8	5	5	6	6	6	78
APL News	7	7	7	10	11	10	6	6	9	6	7	11	8	105*
APL BUG	8	9	8	8	8	11	10	10	8	7	11	8	11	117
Big APL	9	11	10	11	7	7	8	9	10	10	10	7	10	119
BBS\APL	11	8	11	9	10	8	11	11	7	8	8	12	13	127
Ed. Vect.	10	10	13	7	12	9	9	7	11	9	9	10	7	123
APL CAM	12	12	12	13	9	12	12	12	12	12	12	13	12	155*
Les Nouv	13	13	9	12	13	13	13	13	13	13	13	9	9	156*

Row sum ranks are almost identical to Table 2, thus showing statistical robustness — different data slices yield similar results.

VI. Satisfaction with SIGAPL Conferences

Question E asked: *"How satisfied are you with the Conference features below? (1=most; 5=least)."*

Table 5: Satisfaction with SIGAPL Conference Features

Full Sample			most - satisfied - least					
N	%	Conference Feature	1	2	3	4	5	ARS
74	40	Proceedings	41	19	8	5	1	1.73
66	35	Location	36	21	2	3	4	1.76
65	35	Vendor information	28	28	5	2	2	1.80
54	29	Program	20	21	9	1	3	2.00
38	20	Social/Vacation	11	19	4	2	2	2.08
49	26	Workshops	14	17	12	6	0	2.20
48	26	Banquet	17	16	6	5	4	2.23
42	23	Birds of a Feather	11	10	11	8	2	2.52
52	28	Tutorials	11	15	15	7	4	2.58
41	22	Housing	6	17	11	2	5	2.59
39	21	Poster Sessions	4	14	14	5	2	2.67
55	30	Software Exchange	10	13	18	12	2	2.69
75	40	Cost	13	17	17	12	16	3.01
34	18	Business/job oppty	2	6	10	9	7	3.38

Again, a different picture emerges from the split sample.

Table 6: Split Sample: SIGAPL Annual Conferences

SIG				Non-SIG			
N	%		ARS	N	%		ARS
65	54	Proceedings	1.69	5	8	Banquet	1.40
56	46	Location	1.75	4	6	Social/Vacation	1.50
57	47	Vendor information	1.82	8	12	Vendor information	1.63
48	40	Program	1.98	10	15	Location	1.80
34	28	Social/Vacation	2.15	9	14	Proceedings	2.00
44	36	Workshops	2.18	6	9	Program	2.17
43	36	Banquet	2.33	10	15	Cost	2.30
37	31	Birds of a Feather	2.49	3	5	Housing	2.33
47	39	Tutorials	2.51	5	8	Workshops	2.40
38	31	Housing	2.61	6	9	Software Exchange	2.50
36	30	Poster Sessions	2.67	4	6	Business/job oppty	2.50
49	40	Software Exchange	2.71	3	5	Poster Sessions	2.67
65	54	Cost	3.12	5	8	Birds of a Feather	2.80
30	25	Business/job oppty	3.50	5	8	Tutorials	3.20

VI. Nullius in Verba *(Check it yourself)*

This article just scratches the surface. Do your own analysis to expand or correct this article. A complete data set (edited to assure confidentiality) is available: as `SURV94.ZIP` from the BBS\APL at 703-528-7617, a Dallas BBS at 214-682-9656, an Actuary BBS at 908-232-7464, by `ftp` from `watserv1.uwaterloo.ca/languages/apl`, and by mail on disk for $US5 post-paid world-wide from Dick Holt, 3802 N. Richmond Street, Arlington VA 22207 USA. `SURV94.ZIP` contains survey data in multiple ASCII files, and `SURV94.AWS` (v8+) with all data vars and analysis fns.■

References

[1] SIGAPL extends special thanks to Cynthia Rose at ACM, for timely help with ACM demographic data for SIGAPL.

[2] *"Exploratory Data Tables, Trends and Shapes"*, by David Hoaglin, Frederick Mosteller, and John Tukey, John Wiley and Sons, 1985

[3] *"Practical Non-Parametric Statistics"*, by W. J. Conover, John Wiley and Sons, 1980

[4] *"How Much Difference *Makes* a Difference?"*, by Dick Holt, Quote Quad, Vol 25, Number 2 (December 1994)

[5] Bill Chang, private communication, November 1994